

Sustaining the momentum, moving the DataVault project to a service

Claire Knowles (University of Edinburgh) @cgknowles

CC-BY



THE UNIVERSITY
of EDINBURGH



GRAVIRANJE KLJUČAVNIC
tel.: 041 585 648

University of Edinburgh simplified data lifecycle



BEFORE

create a data
management
plan

DURING

working with
data

AFTER

share and
archive your
data

TRAINING
&
SUPPORT

BEFORE

DMPOnline

Worktribe

DURING

Active Data
Store

Electronic
LabNotebooks

HPC

DataSync

DataSafe
Haven

AFTER

DataShare
(Open)

DataVault
(Closed)

DataSync
(Sharing)

PURE

TRAINING AND SUPPORT

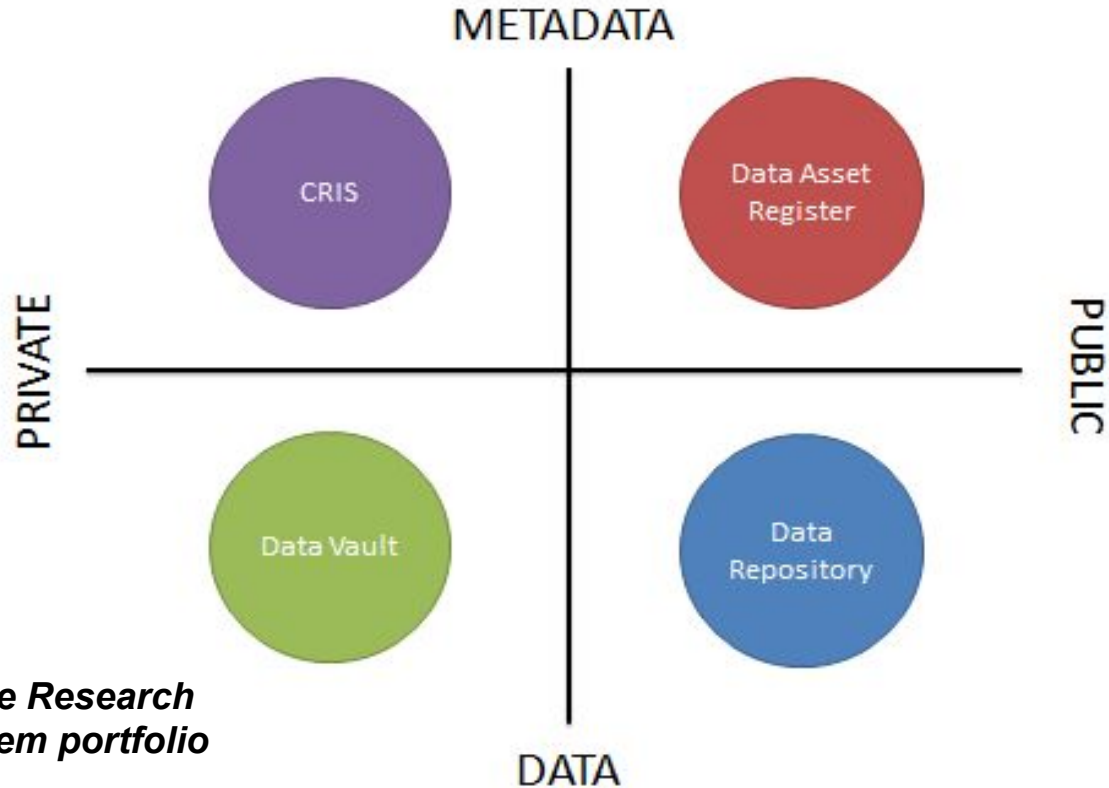
1-2-1s

Workshops

Online Support

MOOC

AFTER



***DataVault within the Research
Data Services System portfolio***

BEFORE Funders Requirements

Alzheimers Society	N	N	N/A		No mention of RDM or data sharing in any of the applicant guidance. Open Access for publications is required
Anatomical Society	N	N	N/A		No mention of RDM or data sharing in any of the applicant guidance.
Arthritis Research UK	N	N	N/A		No mention of RDM or data sharing in any of the applicant guidance.
AXA Research Fund	N	N	N/A		The AXA Research Fund supports the basic principle of Open Access to research data. Nevertheless, considering the fast evolution of current debates and the diversity of approaches, all its funded researchers to follow best practices: * By retaining files of all the research data they have produced and used during the course of their work, and that they be prepared to share these data with other researchers * By seeking to establish and practice good research data management in accordance with the current best practices in their respective fields and to share their data with other researchers
BBSRC	Y	Y	10 years		Researchers are expected to ensure that data are maintained for a period of 10 years after the completion of the research project in suitable accessible formats using established standards and made available on request
BRE Trust	N	N	N/A		No grant applicant guidance available
British Academy	N	N	N/A		No mention of RDM or data sharing in any of the applicant guidance.
British Council	N	N	N/A		No grant applicant guidance available
British Heart Foundation	N	N	N/A		No mention of RDM or data sharing in any of the applicant guidance.
British Society for Neuroendocrinology	N	N	N/A		No mention of RDM or data sharing in any of the applicant guidance.
Cancer Research UK	Y	Y	5 years		Once the funding for a project has ceased researchers should preserve all data resulting from that grant to ensure that data can be used for followup or new studies. We expect researchers to make data available on request
Carnegie Trust	N	N	N/A		No grant applicant guidance available
Carnegie trust for the Universities of Scotland	N	N	N/A		26.The Trust does not currently impose any data-sharing or data-storage requirements on recipients of its Grants, but may reconsider its policy in this respect.
Chest, Heart and Stroke Scotland	N	N	N/A		No grant applicant guidance available
Chief Scientist Office	N	Y	5 years		16. Preserving and Sharing Research Data 16.1 CSO, in common with other public research funders, strongly encourages the sharing of data from research it supports. Where the data may be of interest to researchers, the data should be preserved in a way that encourages other analysts to use them. The best method for ensuring this is to deposit the data with full supporting documentation in a public repository (e.g. 13.8). 16.2 CSO recognises that the original investigator has a right to a limited period of exclusive use of the data, that secondary analyses may be most fruitfully conducted in collaboration with the investigator. 16.3 Whether or not the data are likely to be used for secondary analysis, the Chief Investigator must ensure that the raw data or results are stored for a minimum period of 5 years as indicated in the notice of funding.
CHILDREN with CANCER UK	N	N	N/A		No grant applicant guidance available
Daphne Jackson Trust	N	N	N/A		No mention of RDM or data sharing in any of the applicant guidance.

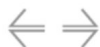


Sustaining Research Data

- Know what data is in the Vault
- One Vault has 1 or more deposits
- Manifest of files is stored in the database and the deposit
- Deposit is in the LOC BagIt standard
- Checksum for each: deposit, chunk, file
- Geographical resilience, 3 copies
 - 1 in Edinburgh (UoE)
 - 1 outer Edinburgh (UoE)
 - 1 in the cloud, within the UK

Easy to use

- Integration with PURE
- Notifications of deposit and retrieval
- Metadata extraction
- Sharing of permissions
- Competitively priced
- Integration with single sign-on (LDAP)



DataVault - Home Page (tab 1 open)



DataVault

Archive your data securely, with automatic encryption

[Storage options](#) | [Contact](#) | [More Info](#) | [Logout](#)

Contact will go to IS helpline

Transfer your data into long term storage and free up your working data store.

Why?:

- You have data that needs to be protected from accidental change or deletion
- Your funder requires you to keep data safe for a set time period
- You are not able to share your data openly, but want to keep a copy that can be requested

Current Vaults

If you already have some vaults they will appear here

Vault name	Deposit	Owner name	Date Created	Review Date	Deposit/Retrieval Completeness
name here	Deposit	name here	00/00/0000	00/00/0000	Deposit <div></div>
name here	Closing to new deposits in less than 3 months	name here	00/00/0000	00/00/0000	Retrieval <div></div>
name here	Closed	name here	00/00/0000	00/00/0000	Retrieval

Create and deposit in a new vault

1. Select Dataset

2. Create Vault

3. Deposit

You must record the details of your dataset in [PURE](#). If you haven't yet created a Pure dataset record containing the details of the data you're about to deposit in the vault, please do so.

Please select the Pure record describing the data that will be contained in this vault from the list below:

Each record will have one of the following next to it:



THE UNIVERSITY
of EDINBURGH

Information Services

RESEARCH DATA SERVICE

Protect data from accidental
change or deletion

Meet funder requirements to
keep data safe for a set time
period

Keep copies of data you are not
able to share openly

Create new vault [?]

You must record the details of your dataset in [PURE](#). If you haven't yet created a Pure dataset record containing the details of the data you're about to deposit in the vault, please do so.

Please select the Pure record describing the data that will be contained in this vault from the list below:

Biographies of individuals mentioned in the papers of Alexander Carmic [?]

The 'owner' /PI will be billed for any deposit, against the Grant ID they provided in relation to the vault, unless they have made other arrangements with Information Services Research Services. Rates are advertised on the Research Services charges page.

You will be aware that the data you place in the Vault belongs to the University of Edinburgh and you will only be able to retrieve the data as long as you are employed by the University of Edinburgh.

Vault Name [?]

Enter a descriptive name for the Vault e.g. the project or experiment.



Description [?]

Retention Policy [?] [Read more about retention policies](#)

Please choose a retention policy



Engagement with researchers

- Research Data Steering Group
- Project Board
- Requirements gathering
- Interim Service use
- Training and support
- User Acceptance Testing
- College Reps on the Project Board
- Roadshows and workshops
- Support and help on the university webpages

<https://www.ed.ac.uk/information-services/research-support/research-data-service/sharing-preserving-data/datavault>



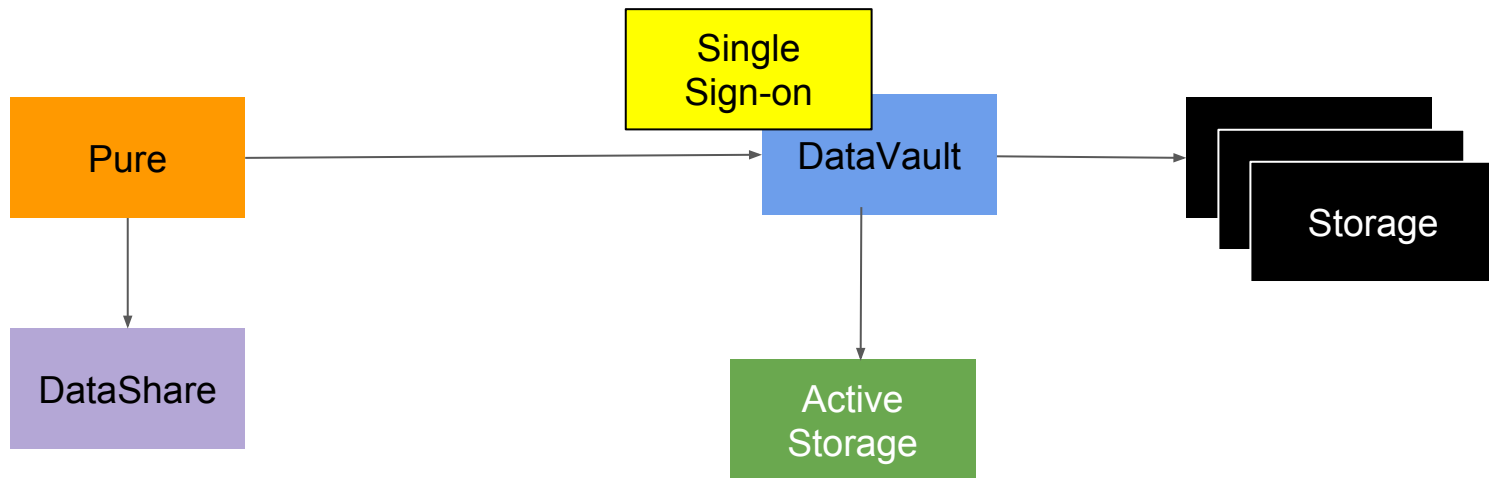
Costing

- Aim to be cheaper than active storage
- Compared costs to cloud storage offerings
- Free storage for first 100GB per user
- Initial cost is £500 per terabyte for 10 years
- Full economic costing is underway

Photo by [Joe Tree](#) on [Unsplash](#)



Integrations



Internal or External?

- **Data curation services are the most likely to be externalized.** Infrastructure and other technical resources needed to support Curation services are expensive to implement and manage and require expertise that may not be readily available on campus. Additionally, Curation services tend to be scalable, and therefore well-suited for provision as a shared service.

<http://www.oclc.org/blog/main/libraries-and-rdm-three-decisions-three-components-three-realities/>

DataVault Architecture

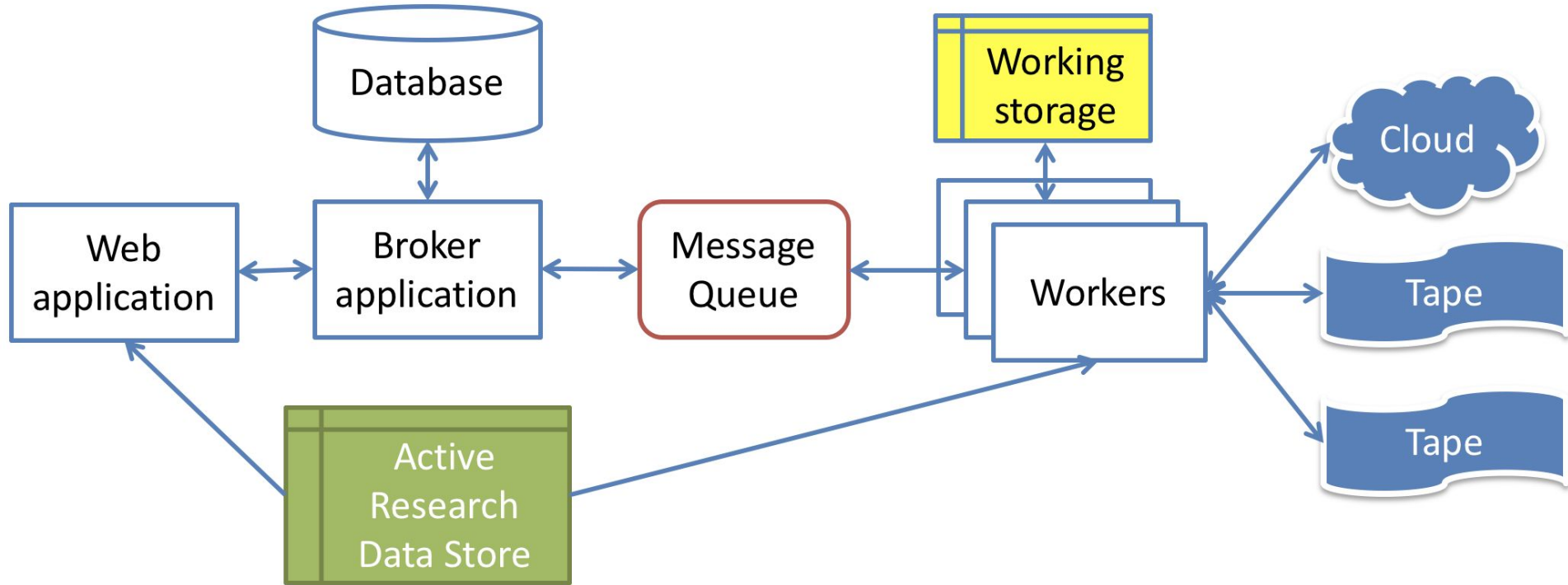




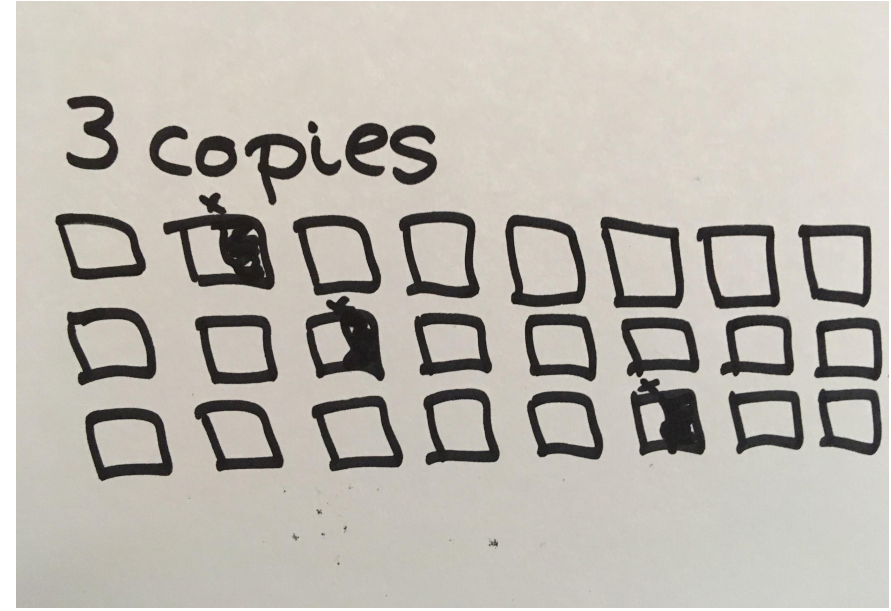
Photo by [Marten Bjork](#) on [Unsplash](#)

Personal Data Considerations

General Data Protection Regulation (GDPR) became law on 25th May 2018

- Research data and GDPR
- Encryption of all deposits
- Encrypt all
- Key management
- Chunking of deposits

Implications of encryption





Communications and Engagement

- Training
- Workshops
- Liaison Librarians
- Research Data Support
- College IT Officers
- Blog posts
- Videos

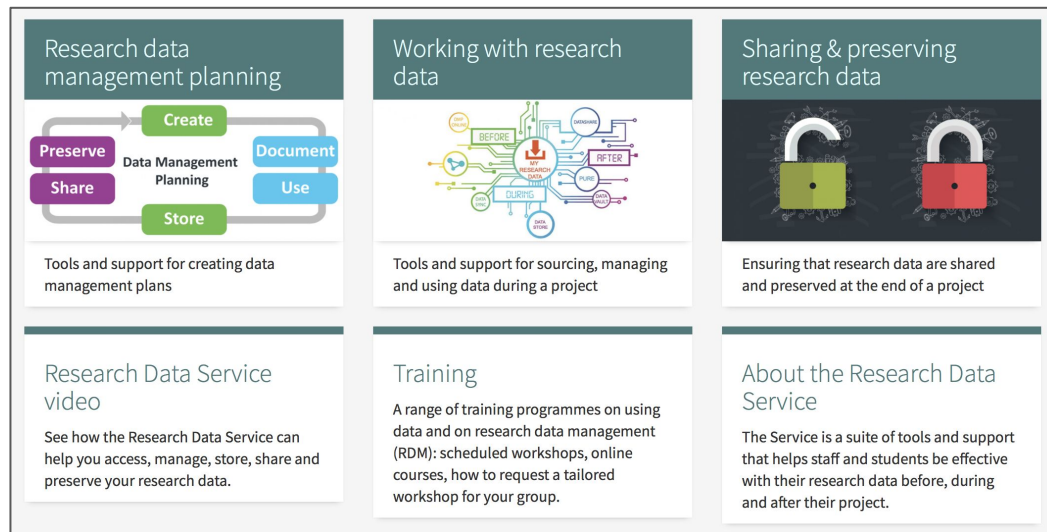


Photo by [Park Troopers](#) on [Unsplash](#)



Is this digital preservation?

No

But ...

- Multiple copies
- Learnt from best practice
- Backlog of research data before active preservation

Thanks

Claire Knowles

@cgknowles

claire.knowles@ed.ac.uk

<https://github.com/DataVault/datavault>

<https://www.ed.ac.uk/information-services/research-support/research-data-service/sharing-preserving-data/datavault>

[https://figshare.com/articles/Sustaining the momentum moving the DataVault project to a service/6453122](https://figshare.com/articles/Sustaining_the_momentum_moving_the_DataVault_project_to_a_service/6453122)